

Weekly Report

12/29/2015 - 01/11/2015

Jing XIA

January 11, 2015

1 Summary

This week I mainly focus on the data inspection project.

2 Projects

2.1 Project 1 - Rank Visualization

-

2.2 Project 2 - Data Inspection

After a second thought of view similarity proposal, we doubt its usefulness in revealing inter-view insights, and discard the idea. We can do nothing with views of similar pattern regardless of the shape of the view scatterplot. Then what do users want to know of the dataset indeed? Correlation. And how can we better explore correlations between views? Instead of correlation coefficient, we use mutual information, which uncovers both linear and non-linear correlations between views.

Highly correlated views have larger mutual information, and vice versa. To further correlate local regions between views, we apply the notion of PMI (point-wise mutual information). The optimization of local sum of PMIs can indicate how two views are highly correlated. Moreover, for views containing categorical dimension(s), such optimization can also be used for sorting. Better sorting clusters highly correlated variables of the underlying categorical dimension(s). Here is how we proceed the new method.

1. Upon initial we compute pairwise view correlation of 1D, 2D, and hybrid views.
2. Then we construct quartet units from all the views based on the correlation.
3. Then we apply the method in “Qualitative Organization of Collections of Shapes via Quartet Analysis” for **degree of separation** of views and layout.

4. For visualization, views of sequential dimensions are sorted sequentially while views of categorical dimensions are sorted with local sum of PMIs optimized. Also, highly correlated local areas between views are highlighted pairwise.
5. For interaction, users can brush on variables or lasso value areas in the views to filter the dataset progressively.

We also would like to distinguish our method with what's described in the paper "An Information-Aware Framework for Exploring Multivariate Data Sets". They did clustering based on the pre-defined correlations. Users can brush on PCP based on both variable values and predefined values related to mutual information. In terms of our method, we have 3 major functions that distinguishes. First, we compute correlations of 1D views, 2D views and hybrid views. Second, we define and optimize local sum of PMIs to locate highly correlated local areas between views. And last, such local optimization can also be used to assist variable sorting in categorical views.

2.3 Project 3 - NBA Game Visualization

-

3 Paper Reading

-

4 Miscellaneous

Load 911 dataset into Mountdb.

5 To Do List

1. Data inspection project — test the mutual information methods.
2. NBA project discussion and data inspection project discussion.

References